

Effect of the target positions on the rapid identification of aluminum alloys by using filament-induced breakdown spectroscopy combined with machine learning

Xiaoguang Li(李晓光), Xuetong Lu(陆雪童), Yong Zhang(张勇),Shaozhong Song(宋少忠), Zuoqiang Hao(郝作强), and Xun Gao(高勋)

Citation: Chin. Phys. B, 2022, 31 (**5**): 054212. DOI: 10.1088/1674-1056/ac3810
Journal homepage: http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

Interrogation of optical Ramsey spectrum and stability study of an ^{87}Sr optical lattice clock

Jing-Jing Xia(夏京京), Xiao-Tong Lu(卢晓同), and Hong Chang(常宏)
Chin. Phys. B, 2022, 31 (**3**): 034209. DOI: 10.1088/1674-1056/ac11e0

An approach to gas sensors based on tunable diode laser incomplete saturated absorption spectra

Wei Nie(聂伟), Zhen-Yu Xu(许振宇), Rui-Feng Kan(阚瑞峰), Mei-Rong Dong(董美蓉), and Ji-Dong Lu(陆继东)
Chin. Phys. B, 2021, 30 (**6**): 064213. DOI: 10.1088/1674-1056/abd769

A pressure-calibration method of wavelength modulation spectroscopy in sealed microbial growth environment

Kun-Yang Wang(王坤阳), Jie Shao(邵杰), Li-Gang Shao(邵李刚), Jia-Jin Chen(陈家金), Gui-Shi Wang(王贵师), Kun Liu(刘琨), and Xiao-Ming Gao(高晓明)
Chin. Phys. B, 2021, 30 (**5**): 054203. DOI: 10.1088/1674-1056/abda32

Improve the performance of interferometer with ultra-cold atoms

Xiangyu Dong(董翔宇), Shengjie Jin(金圣杰), Hongmian Shui(税鸿冕), Peng Peng(彭鹏), and Xiaoji Zhou(周小计)
Chin. Phys. B, 2021, 30 (**1**): 014210. DOI: 10.1088/1674-1056/abcf33

Effect of the distance between focusing lens and target surface on quantitative analysis of Mn element in aluminum alloys by using filament-induced breakdown spectroscopy

Xue-Tong Lu(陆雪童), Shang-Yong Zhao(赵上勇), Xun Gao(高勋), Kai-Min Guo(郭凯敏), and Jing-Quan Lin(林景全)
Chin. Phys. B, 2020, 29 (**12**): 124209. DOI: 10.1088/1674-1056/abb3ef

---

Effect of the target positions on the rapid identification of aluminum alloys by using filament-induced breakdown spectroscopy combined with machine learning

Xiaoguang Li(李晓光)[1], Xuetong Lu(陆雪童)[2], Yong Zhang(张勇)[1], Shaozhong Song(宋少忠)[3,†], Zuoqiang Hao(郝作强)[4], and Xun Gao(高勋)[2,‡]

[1] *School of Electrical Information, Changchun Guanghua University, Changchun* 130033, *China*
[2] *School of Physics, Changchun University of Technology, Changchun* 130600, *China*
[3] *School of Information Engineering, Jilin Engineering Normal University, Changchun* 130052, *China*
[4] *School of Physics and Electronic Sciences, Shandong Normal University, Jinan* 250358, *China*

(Received 22 August 2021; revised manuscript received 6 November 2021; accepted manuscript online 10 November 2021)

Filament-induced breakdown spectroscopy (FIBS) combined with machine learning algorithms was used to identify five aluminum alloys. To study the effect of the distance between focusing lens and target surface on the identification accuracy of aluminum alloys, principal component analysis (PCA) combined with support vector machine (SVM) and K-nearest neighbor (KNN) was used. The intensity and intensity ratio of fifteen lines of six elements (Fe, Si, Mg, Cu, Zn, and Mn) in the FIBS spectrum were selected. The distances between the focusing lens and the target surface in the pre-filament, filament, and post-filament were 958 mm, 976 mm, and 1000 mm, respectively. The source data set was fifteen spectral line intensity ratios, and the cumulative interpretation rates of PC1, PC2, and PC3 were 97.22%, 98.17%, and 95.31%, respectively. The first three PCs obtained by PCA were the input variables of SVM and KNN. The identification accuracy of the different positions of focusing lens and target surface was obtained, and the identification accuracy of SVM and KNN in the filament was 100% and 90%, respectively. The source data set of the filament was obtained by PCA for the first three PCs, which were randomly selected as the training set and test set of SVM and KNN in 3:2. The identification accuracy of SVM and KNN was 97.5% and 92.5%, respectively. The research results can provide a reference for the identification of aluminum alloys by FIBS.

Keywords: filament-induced breakdown spectroscopy (FIBS), principal component analysis (PCA), support vector machine (SVM), K-nearest neighbor (KNN), aluminum alloys identification

PACS: 42.62.Fi, 52.38.Hb **DOI:** 10.1088/1674-1056/ac3810

1. Introduction

Laser-induced breakdown spectroscopy (LIBS) is an elemental analysis method. In qualitative and quantitative analysis of the constituent elements of substances (solid, liquid or gas samples), the emission spectrum of plasma is generated by ablation of target surface by a tightly focused nanosecond laser.[1,2] Due to changes in the interaction between laser and matter, femtosecond LIBS (fs-LIBS) can improve the parameters of LIBS quantitative analysis (for example, better reproducibility, reduced matrix effect and reduced damage).[3,4] Other advantages of fs-LIBS include low ablation threshold, signal accumulation of a large number of pulses, and reduced heat-affected zone.[5,6] Due to the dynamic balance between optical Kerr self-focusing and plasma de-focusing, the femtosecond laser beam is focused in the air to form a long filament with the clamped laser intensity of 10^{13} W/cm^2–10^{14} W/cm^2.[7] Compared with traditional long-pulse laser-induced breakdown spectroscopy, a significant advantage of FIBS is that it can overcome diffraction to transmit high-intensity filaments over long distances. It has been widely used in quantitative analysis of substances, detection of explosive substances, heavy metal pollution in water, biological fields, isotope detection, environmental detection, and other fields.[8–14]

In recent years, FIBS combined with machine learning algorithms to detect and identify samples has attracted much attention. Kalam *et al*.[15] used fs-LIBS and principal component analysis (PCA) to detect and identify explosive molecules (nitroimidazole and nitropyrazole). The Schmidt–Cassegrain telescope collects radiant plasma at about 8.5 m. The focal length of the focusing lens is 10 cm, 30 cm, 50 cm, 100 cm, and 200 cm. As the focal length of the lens increases, the contribution rate of the first three PCs drops from 99% to 52%. Kalam *et al*.[16] used FIBS combined with PCA to detect and identify metals, alloys, and bimetals at a distance (6.5 m). When the ICCD gain is 1000 and 2500, the first three PCs of metal and bimetal account for 73% and 98% of the total variance. Junjuri *et al*.[17] used fs-LIBS combined with PCA and artificial neural networks (ANN) to obtain a 100% correct prediction rate for post-consumer plastic identification.

†Corresponding author. E-mail: songsz@jlenu.edu.cn
‡Corresponding author. E-mail: lasercust@163.com

© 2022 Chinese Physical Society and IOP Publishing Ltd

Fu *et al.*[14] used FIBS combined with PCA to identify tree species (willow, pine, and poplar) and their growth environment, and the first three PCs accounted for 99.83% of the total data. Narla *et al.*[18] used FIBS to extract features from PCA as input variables for support vector machine (SVM) to identify metals and metal alloys. Both the color CCD image and femtosecond LIBS achieve 100% identification accuracy. Kalam *et al.*[19] used fs-LIBS and FIBS to identify four types of rocks by using PCA in near-field and standoff modes (15 cm, 6.5 m), and the first three PCs accounted for 93%, 82%, 97%, and 91% of the total variance in the data set in the nearfield and standoff mode, respectively.

Different laser pulses (such as laser wavelength, pulse duration, and laser energy) and experimental conditions (for example, the direction of incident laser to target surface, the method of collecting spectral signals, or the distance between focusing lens and target surface) affect the spectral emission of laser-induced plasma.[20,21] Harilal *et al.*[22] found that the spectral radiation intensity and spatial distribution of the electron density of copper ablated by filament have the same trend, which firstly increases and then decreases along the filament. The variation of the spectral radiation intensity at different positions of the filament is related to the atomic density, which affects the ablation efficiency. Increasing the filament laser pulse energy, the filament plasma electron density also becomes higher, because there is an energy reservoir around the filament. Yao *et al.*[23] studied the influence of femtosecond laser energy on ablation characteristics. When the femtosecond laser energy increases, the internal energy of the energy reservoir around the filament also increases. The different distances between the focusing lens and target surface will result in different spot sizes on the target surface, which will affect the interaction between laser and target. This results in changes in laser-induced plasma spectral radiation characteristics, ablation efficiency, and FIBS analysis performance.[24,25] In our previous work,[26] we studied the effect of the distance between focusing lens and target surface on quantitative analysis of Mn in aluminum alloys by using FIBS. It is found that the distance between the focusing lens and target surface affects spectral stability and accuracy of quantitative analysis, and the filament is better than the pre- and post-filaments. Therefore, this paper uses FIBS combined with machine algorithms to identify aluminum alloys, uses PCA to analyze the data set of intensity and intensity ratio of fifteen characteristic spectral lines of FIBS spectrum, and combines SVM and K-nearest neighbor (KNN) to analyze the effect of the distance between the focusing lens and target surface on identification accuracy of aluminum alloys.

2. Experiments and samples

Figure 1 is an experimental setup for identifying aluminum alloys by using FIBS. Ti: sapphire laser system (Coherent, repetitive frequency of 1 kHz, wavelength of 800 nm) generates 45 fs laser pulses. The femtosecond laser beam is focused on aluminum alloys by a K9 lens 1 with the focal length of 1000 mm. The plasma emission spectral signal is collected by lens 2 (BK7, $f = 75$ mm, diameter 50 mm) and coupled to Andor Echelle 5000 spectrometer (resolution of 0.05 nm at 500 nm, and detection spectrum range of 200 nm–880 nm) equipped with ICCD through an optical fiber (diameter 600 μm). The aluminum alloy target is placed on a three-dimensional (3D) translation stage in order to avoid excessive ablation. To obtain FIBS spectrum in accumulation mode (5 times accumulation), delay time, and the gate width are set 4 μs and 10 μs, respectively. A one-dimensional (1D) translation stage is used to change the distance between lens 1 and the target surface, and the forming filament is about 40 mm (observed by the naked eye). The whole experiments are done at the experimental conditions of pulse energy of 2 mJ, standard atmospheric pressure, room temperature of 22 °C, and relative humidity of 25%.

In our experiment, five kinds of aluminum alloy samples (1060/2024/5052/6061/7075) were purchased on the market with length of 5 cm, the width of 5 cm, and thickness of 2 mm. Table 1 is the element concentrations of aluminum alloy samples.

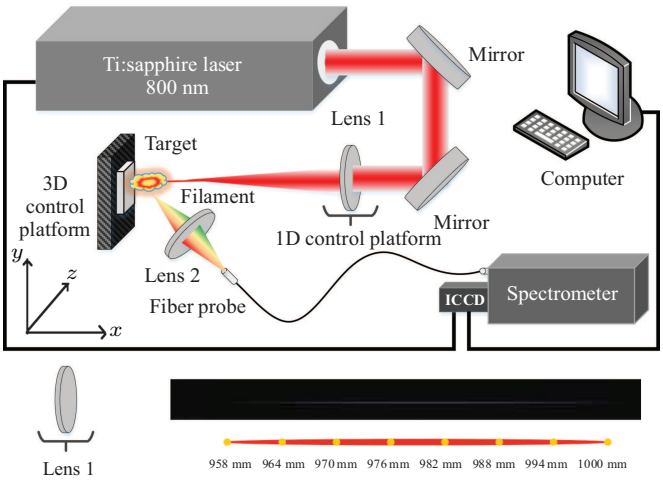


Fig. 1. FIBS system setup.

Table 1. Element concentrations of aluminum alloy samples (in units of wt%).

Class	Si	Fe	Cu	Mn	Mg	Zn
1060	0.25	0.35	0.05	0.03	0.03	0.05
2024	0.5	0.5	3.8–4.9	0.3–0.9	1.2–1.8	0.25
5052	0.25	0.4	0.1	0.1	2.2–2.8	0.1
6061	0.4–0.8	0.7	0.15–0.4	0.15	0.8–1.2	0.25
7075	0.4	0.5	1.2–2.0	0.3	2.1–2.9	5.1–6.1

3. Data processing

3.1. PCA

PCA is an unsupervised data dimensionality reduction method. It calculates the eigenvector of data covariance matrix, and projects each variable onto largest eigenvector to reduce the number of variables and ensure that each variable is independent.[27–30] When using PCA, for original data $X_{m\times n}$ [m represents the number of samples (number of spectra), n represents the number of features (wavelength)], standard conversions of samples are

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \tag{1}$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2, \tag{2}$$

and the feature covariance matrix is

$$\mathrm{cov}(x_1, x_2, \ldots x_n) = \begin{pmatrix} \mathrm{cov}(x_1,x_2) & \cdots & \mathrm{cov}(x_1,x_n) \\ \vdots & \ddots & \vdots \\ \mathrm{cov}(x_n,x_1) & \cdots & \mathrm{cov}(x_n,x_n) \end{pmatrix}. \tag{3}$$

Find eigenvalues and eigenvectors of covariance matrix

$$xR = \omega R, \tag{4}$$

where eigenvalue ω is the diagonal element, and R is eigenvector of x corresponding to the eigenvalue ω. The principal component is selected according to the contribution rate of the principal component. The principal component (PC) represents the variance of the data set, and PC1 is the highest.

3.2. SVM

SVM establishes the optimal hyperplane based on the principle of maximizing the interval.[31–33] The training sample set is (x_i, y_i), x is the sample attribute value, and y is the result label. A linear function is constructed in the high-dimensional space to realize the nonlinear discriminant function in the low-dimensional space, and the optimal linear decision function analysis data is obtained. A hyperplane $f(x) = \omega^{\mathrm{T}} x + b$ is established in the data space to find the optimal classification hyperplane. This problem is solved by the Lagrangian method, and the dual form is obtained by

$$\left(\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j \alpha_i \alpha_j (x_i . x_j) - \sum_{i=1}^{n} \alpha_i\right), \tag{5}$$

$$0 \le \alpha_i \le C, \quad i = 1, 2 \ldots, n, \tag{6}$$

$$\sum_{i=1}^{n} \alpha_i y_i = 0. \tag{7}$$

The decision function is

$$f(x) = \mathrm{sign}\left(\sum_{i=1}^{n} \alpha_i y_i K(x_i, x_j) + b\right). \tag{8}$$

Select the kernel function as the Gaussian radial basis function (RBF)

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right). \tag{9}$$

Finally, the decision function is

$$f(x) = \mathrm{sign}\left(\sum_{i=1}^{n} \alpha_i y_i \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) + b\right), \tag{10}$$

where σ represents the width of the Gaussian kernel function.

3.3. KNN

KNN is a classic pattern classification method. The three factors are the selection of K value, distance measurement method, and classification decision rules.[34–36] Generally, the appropriate K value is chosen according to the distribution of the sample. KNN uses simple distances, such as Euclidean distance $\sqrt{\sum_{i=1}^{K} (X_i - Y_i)^2}$ to measure the difference between input vectors and samples. KNN usually uses a majority voting method for classification prediction. That is to say, the K samples closest to the prediction sample in the training set, and the category with the largest number of categories in the prediction. Suppose the training set is

$$\begin{aligned} T = {} & \{(x_1, y_1), (x_2, y_2), \ldots, (x_{i-1}, y_{i-1}), (x_i, y_i)\}, \\ & (i = 1, 2, \ldots, N), \end{aligned} \tag{11}$$

where X is the test set, T is the training set, N is the number of samples, and m is the category of X. The distance $d(X, T)$ between the test set X and the training set T is calculated. The distance measurement methods include Euclidean distance, Manhattan distance, etc. The distance between X_i and X_j is

$$L(X_i, X_j) = \left(\sum_{l=1}^{n} \left|X_i^l - X_j^l\right|^P\right)^{\frac{1}{P}}, \tag{12}$$

where $x_i, x_j \in R^n$, $P \ge 1$. When $P = 1$, it is the Manhattan distance, and when $P = 2$ it is the Euclidean distance. The classes of test set X are

$$m_x = \mathrm{argmax} \sum_{j \in b} d(X_i = T_i), \tag{13}$$

where d is an indicator function. When $X_i = T_i$, $d = 1$, otherwise $d = 0$.

4. Results and discussion

4.1. FIBS spectra

Figure 2 shows the FIBS spectra of five aluminum alloys. At the distance of 976 mm between the focusing lens and target surface, the 20 sets of spectral data were averaged and normalized. It can be seen that the AlO radical $B^2\Sigma^+ X^2\Sigma^+$ transition spectrum line is detected in the wavelength range of 440 nm–540 nm. This is due to the coexistence of several reactions among the excited radicals, atoms, molecules, and the environment in plasma generated by the laser, which leads

to the formation or depletion of substances.[37,38] The wavelength range of 550 nm–880 nm has a wavy continuous spectrum, because the Echelle spectrometer diffracts the spectrum in this wavelength range.

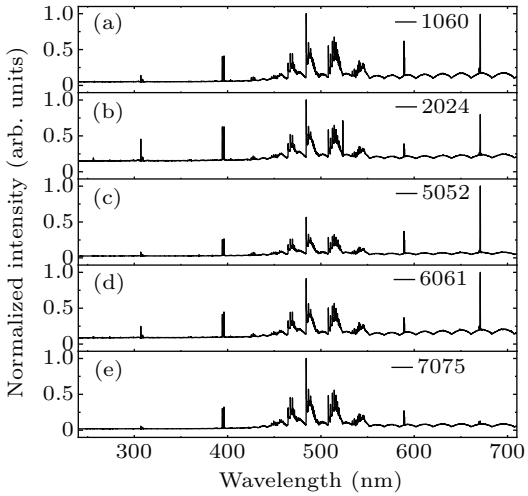


Fig. 2. FIBS spectra of the aluminum alloys: (a) 1060, (b) 2024, (c) 5052, (d) 6061, and (e) 7075.

According to the National Institute of Standards and Technology (NIST) standards database, the FIBS spectrum in the wavelength range of 200 nm–880 nm is marked, and the plasma optical emissions of six elements iron (Fe), silicon (Si), magnesium (Mg), copper (Cu), zinc (Zn) and manganese (Mn) are found. The concentration of elements in experimental samples of aluminum alloy is different, and the intensity of the FIBS characteristic spectrum is also different. The aluminum alloys are identified by the intensity of characteristic spectral lines. The selection of the characteristic spectral line satisfies the high intensity of the spectral line and high signal-to-background ratios (SBR), and there is no strong interference spectral line around the characteristic spectral line. The characteristic spectrum selects fifteen spectrum lines of six elements with an error of 0.1 nm, as shown in Table 2.

Table 2. The 15 characteristic spectral lines.

Element	Emission lines (nm)
Si	251.61, 252.41
Fe	344.06, 587.84,670.96
Cu	282.43, 324.75
Mn	293.93, 348.29, 403.31,
Mg	277.98, 279.79, 285.21
Zn	257.06

4.2. Different positions of filament

The initial power $P_{\rm in}$ of the femtosecond laser pulse is higher than the critical power of self-focusing $P_{\rm cr}$, which will produce filaments. According to the semi-empirical Marburger formula, the position $L_{\rm c}$ where the self-focusing collapse occurs (where the beam radius is the smallest along the propagation direction) is[39]

$$L_{\rm c}=\frac{0.367kw_0^2}{\left\{\left[\left(\frac{P_{\rm in}}{P_{\rm cr}}\right)^{\frac{1}{2}}-0.852\right]^2-0.0219\right\}^{\frac{1}{2}}}. \tag{14}$$

For Gaussian beams, the critical power of self-focusing is $P_{\rm cr}=\frac{3.77\lambda_0^2}{8\pi n_0 n_2}=10$ GW, n_0 is the linear refractive index, n_2 is the nonlinear refractive index, λ is the center wavelength of the laser, k is the wave number of the laser, and the beam radius w_0 is 2.5 mm. The peak power $P_{\rm in}$ of the incident pulse with a pulse width of 45 fs is 4×10^{10} W. The position d where the filament starts to form is[24]

$$d=\frac{f\times L_{\rm c}}{f+L_{\rm c}}, \tag{15}$$

where f is the geometric focal length of the lens, and the calculated $d=966$ mm. The position where the filaments start to form (966 mm) is before the geometric focus (1000 mm) of the focusing lens, because Kerr nonlinear self-focusing focuses the laser beam in advance.[40] According to changing trend of the spectral line intensity, it can be considered that the filament ends at a distance of 992 mm between the focusing lens and target surface. We are studying three interval ranges: (i) the pre-filament area (958 mm–966 mm); (ii) the filament area (966 mm–992 mm); and (iii) the post-filament area (992 mm–1000 mm).

4.3. Result of PCA

The intensities of fifteen characteristic lines of six elements are used as the PCA data set to obtain the PC contribution of the FIBS spectrum. The optimal number of PCs depends on many factors, such as the structure of the original data, the application, and the experience and expertise of the researchers (select the spectrum according to the sample element). Anabitarte *et al.*[41] used PCA combined with SVM to identify nucleoli and analyzed the influence of the number of PCs on identification accuracy. It shows that at least three PCs can obtain accurate results, that is, the cumulative contribution rate of three PCs should be greater than 85%. In the pre-filament area, the filament area and the post-filament area (the distance between the focusing lens and target surface is 958 mm, 976 mm, and 1000 mm, respectively). The first three PCs explained 69.48%, 79.46%, and 64.54% of the total data, and the first three PCs cumulatively explained 83.26%, 89.06%, and 81.57% of the total data. The first three PCs cannot display the total variance of the data set. PCA results are related to experimental fluctuations (instability and limitations of the detection system, unevenness of local samples, etc.) and the inherent characteristics of the characteristic spectrum. The spectrometer (resolution = 0.05 nm) used in our experiment cannot detect certain signals of trace elements.[29]

In order to improve the PCA results, the intensity ratio between different spectral lines is calculated.[14,19] Including Si I 251.61 nm, Si I 252.41 nm, Zn I 257.07 nm, Mg I 277.98 nm, Mg II 279.79 nm, Cu I 282.43 nm, Mg I 285.21 nm, Mn II 293.93 nm, Cu I 324.75 nm, Fe I 344.06 nm, Mn II 348.29 nm, Mn I 403.31 nm, Fe I 587.84 nm, and Fe II 670.96 nm. Because the intensity ratio produces a larger size. That is, Mg I 277.98 nm/Mn II 293.93 nm, Mg II 279.79 nm/Cu I 282.43 nm, Cu I 324.75 nm/Mn II 348.29 nm, Fe I 344.06 nm/Mn I 403.31 nm, Mg I 285.21 nm/Fe II 670.96 nm, Si I 251.61 nm/Zn I 257.07 nm, Zn I 257.06 nm/Fe I 344.06 nm, Si I 252.41 nm/Fe I 587.84 nm, Si I 252.41 nm/Mg II 279.79 nm, Si I 251.6 nm/Mn II 348.29 nm, Si I 251.6 nm/Cu I 324.75 nm, Zn I 257.07 nm/Cu I 282.43 nm, Zn I 257.07 nm/Mn II 299.93 nm, and Zn I 257.07 nm/Mg I 277.98 nm. The calculated spectral signal intensity ratio is the data set of PCA.

Figure 3 is the result of PCA intensities of fifteen characteristic lines in FIBS spectra of five aluminum alloys in the pre-filament area, the filament area and the post-filament area (the distance between the focusing lens and target surface is 958 mm, 976 mm, and 1000 mm, respectively). PCA converts the FIBS spectra of five samples into three PCs. The cumulative interpretation rate of the three PCs by using intensity ratio as the source data set is 97.22%, 98.17%, and 95.31%, respectively, which is better than the PCA results of fifteen spectral line intensities as feature vectors (83.26%, 89.06%, and 81.57%). Because the signal intensity ratio is used as the PCA data set, the result is independent of the system measurement error, which affects the measured spectral signal intensity.

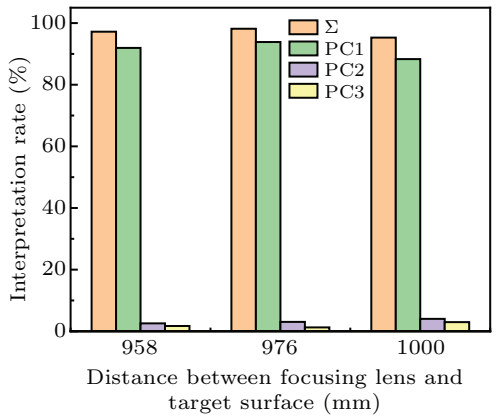


Fig. 3. PCA results at three positions (pre-filament at 958 mm, filament at 976 mm, and post-filament at 1000 mm, respectively).

4.4. Results of SVM/KNN combined PCA

Five aluminum alloy samples have a total of 800 sets of FIBS spectra, and 15 characteristic spectral line intensity ratios are used as a data set for PCA. In the pre-filament area, the filament area and the post-filament area (the distance between the focusing lens and target surface is 958 mm, 976 mm, and 1000 mm, respectively). The cumulative interpretation rates of the first three PCs are 97.22%, 98.17%, and 95.31%, respectively. PC1, PC2, and PC3 replace FIBS spectra to construct feature space vectors, which are used as input variables of SVM and KNN to identify five aluminum alloys. The MATLAB2018 toolbox is used as a data processing tool for 5-fold cross-validation. The confusion matrix represents the identification accuracy of SVM and KNN, and the identification accuracy is the evaluation index for identification.[42] Figure 4 shows the identification results of PCA-SVM and PCA-KNN at three positions (the distance between the focusing lens and target surface is 958 mm, 976 mm, and 1000 mm, respectively). Figure 5 shows the spatial distribution of aluminum alloys identification accuracy. The result of each position is an average of five sample identification results. The identification accuracy of PCA-SVM exceeds 82%, and the identification accuracy of PCA-KNN exceeds 66%. In the three positions, the SVM identification accuracy is 83.4%, 100%, and 82%, respectively. The identification accuracy of KNN is 69.2%, 90%, and 66%, respectively. It can be seen that the identification results of the two algorithms have the same trend, and both rise first and then fall. The identification accuracy of the filament area is higher than that of the pre- and post-filament areas and is stable, and the identification accuracy of SVM is higher than that of KNN.

In our previous work, we found that the quantitative analysis accuracy and spectral stability of the filament area are better than the pre- and post-filament areas.[26] The intensity clamping effect of the filament leads to better laser power stability, sensitivity, and radiation plasma stability in the filament area. The filament carries a small part of the laser energy, and most of the energy is released into the energy reservoir around the filament. The energy store continuously compensates for the energy loss in the ablation process of aluminum alloys, and plays a role in auxiliary ablation. Therefore, the ablation efficiency of filament depends on both the intensity clamping effect in filament and the energy reservoir around the filament.[23,43,44] Before the beam collapse occurs, the Kerr self-focusing effect causes the modulation instability of the laser, which makes the laser intensity uneven. The laser beam is in a divergent state in the post-filament area, the diffraction and dispersion are stronger than the self-focusing effect, and the spectral stability is lower than the pre-filament area. Therefore, the identification accuracy of the pre-filament area is higher than the identification accuracy of the post-filament area.

In filament area (the distance between the focusing lens and target surface is 976 mm), 100 groups of FIBS spectrum data of PCA are randomly selected according to 3:2. The LIBS spectral data of the training set and test set are 60 groups and 40 groups, respectively. SVM model established in the training set predicts 40 sets of spectral data in the test set. The result is shown in Fig. 6, where "*" represents

(a)

Actual class \ Distance/mm	Predicted class: 958					976					1000				
	1060	2024	5052	6061	7075	1060	2024	5052	6061	7075	1060	2024	5052	6061	7075
1060	100					100					60	10	20		10
2024		50		50			100					100			
5052			100					100					100		
6061				100					100			50		50	
7075				33	67					100					100

(b)

Actual class \ Distance/mm	Predicted class: 958					976					1000				
	1060	2024	5052	6061	7075	1060	2024	5052	6061	7075	1060	2024	5052	6061	7075
1060	80				20	100					80	20			
2024		33		67			100					17	17	33	33
5052			67	17	17			100					67	33	50
6061			17	83					100					83	17
7075				17	83		50			50				17	83

Fig. 4. Three positions identification results (pre-filament at 958 mm, filament at 976 mm, and post-filament at 1000 mm, respectively). (a) PCA-SVM and (b) PCA-KNN.

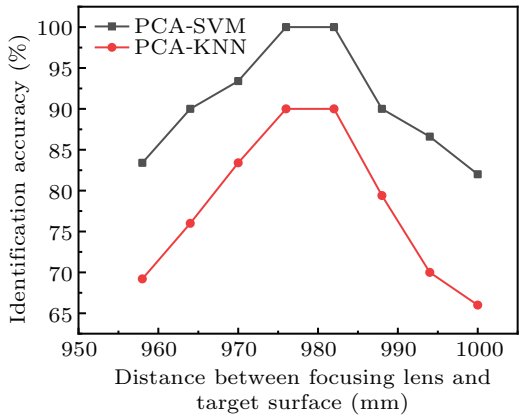


Fig. 5. Spatial distribution of aluminum alloys identification accuracy.

the predicted label, and "○" represents the actual label. The results show that the identification accuracy of PCA-SVM is 97.5%. In the 40 test set spectra, one 7075 was misjudged as 2024. The identification accuracy of PCA-KNN is 92.5%. Two 7075 were misjudged as 2024, and one 1060 was misjudged as 6061. The running time of SVM is 0.62 s–1.74 s, the running time of KNN is 0.37 s–1.39 s, and the running time of KNN is faster than SVM. SVM has certain advantages in processing nonlinear data. The SVM classifier is very simple, and only a small amount of training data is needed to establish a stable classification model. Even if the sample dimension is very high, it will not bring inconvenience to storage and calculation. More stable and accurate identification results can be obtained. The KNN algorithm uses all samples in identification, and needs to be compared with the rest of the points in the set each time. So compared to KNN, SVM has better identification results and stability and is simpler to implement.[45]

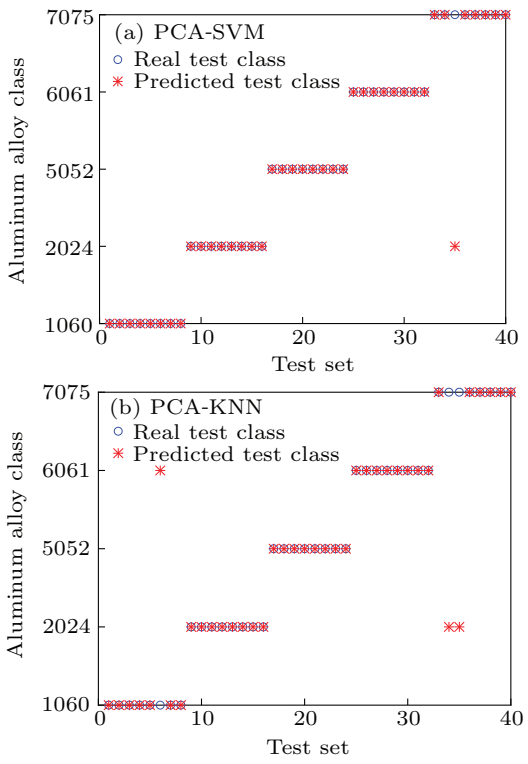


Fig. 6. Spectral identification results in the test set: (a) PCA-SVM and (b) PCA-KNN.

5. Conclusions

This study is based on FIBS, using PCA combined with SVM and KNN to identify aluminum alloys. Two spectral feature extraction methods are proposed for PCA: The characteristic parameter is constituted by the ratio of the intensity of the characteristic spectral line and the intensity of the element spectral line. SVM and KNN are used to identify aluminum alloys. The identification accuracy of the filament area is higher than pre- and post-filament area. The identification accuracy of SVM in the filament area is 100%, and the KNN is 90%. It is because of the clamping effect of the filament and the role of the energy reservoir around the filament. FIBS combines machine learning algorithms to distinguish aluminum alloys, which is beneficial to accurately and quickly detect and identify aluminum alloys during manufacturing, inspection, and recycling processes, and improve the utilization rate of aluminum alloys. In addition, the purpose of this experiment is not to use as many aluminum alloys as possible to prove that this method is applicable to all categories, but to provide a useful method to prove that FIBS can effectively identify aluminum alloy samples. The method of this experiment is also suitable for detecting other substances. For example, FIBS can quickly detect and identify substances from the standoff mode, in dangerous and polluted environments.

Acknowledgement

Project supported by the Natural Science Foundation of Jilin Province, China (Grant No. 20201223 48JC).

References

[1] Noll R, Fricke-Begemann C, Connemann S, Meinhardt C and Sturm V 2018 *J. Anal. At. Spectrom.* **33** 945
[2] Chen C T, Banaru D, Sarnet T and Hermann J 2018 *Spectrochim. Acta, Part B* **150** 77
[3] Byram C, Moram S S B, Shaik A K and Soma V R 2017 *Chem. Phys. Lett.* **685** 103
[4] Rethfeld B, Ivanov D S, Garcia M E and Anisimov S I 2017 *J. Phys. D: Appl. Phys.* **50** 193001
[5] Labutin T A, Lednev V N, Ilyin A A, *et al.* 2015 *Journal of Analytical Atomic Spectrometry* **31** 90
[6] Freeman J R, Harilal S S, Diwakar P K, Verhoff B and Hassanein A 2013 *Spectrochim. Acta, Part B* **87** 43
[7] Deng Y P, Zhu J B, Ji Z G, Liu J S and Chin S L 2006 *Opt. Lett.* **31** 546
[8] Zhao S Y, Gao X, Song X W, Yao S and Lin J Q 2019 *OSA Continuum* **2** 116
[9] Gao X, Du C, Li C, Liu L, Song C, Hao Z Q and Lin J Q 2014 *Acta Phys. Sin.* **63** 095203 (in Chinese)
[10] Abdul K S And Venugopal R S 2019 *OSA Continuum* **2** 554
[11] Li H L,Zang H W,Xn H L,Sun H B, Baltuška A and Polynkin P 2019 *Glob Chall* **3** 1800070
[12] Xu H L, Liu W and Chin S L 2006 *Opt. Lett.* **31** 1540
[13] Hartig K C, Ghebregziabher I and Jovanovic I 2017 *Sci. Rep.* **7** 43852
[14] Fu Y, Hou M Y, Zang H W, Li H, Chin S L and Xu H 2019 *Spectrochim. Acta, Part B* **155** 107
[15] Abdul K S, Nageswara R E, Hamad S and Byram C 2018 *Opt. Express* **26** 8069
[16] Abdul K S, Ajmathulla and Venugopal R S. 2018 *Opt. Lett.* **43** 3465
[17] Rajendhar J and Manoj K G 2019 *J. Anal. At. Spectrom.* **34** 1683
[18] Narla L M and Rao S V 2020 *Appl. Phys. B* **126** 18
[19] Kalam S A, Rao S V B M, Jayananda M and Rao S V 2020 *J. Anal. At. Spectrom.* **35** 3007
[20] Wang W, Sun L X, Wang G D, Zhang P and Dong W 2020 *J. Anal. At. Spectrom.* **35** 357
[21] Harilal S S, Diwakar P K, Polek M P and Phillips M C 2015 *Opt. Express* **23** 15608
[22] Harilal S S, Yeak J and Phillips M C 2015 *Opt. Express* **23** 27113
[23] Yao S, Zhang J, Gao X, Zhao S Y and Lin J Q 2018 *Opt. Commun.* **425** 152
[24] Xu W P, Chen A M, Wang Q Y, Zhang D, Wang Y, Li S Y, Jiang Y F and Jin M X 2019 *J. Anal. At. Spectrom.* **34** 1018
[25] Harilal S S, Yeak J, Brumfield B E and Phillips M C 2016 *Opt. Express* **24** 17941
[26] Lu X T, Zhao S Y, Gao X, Guo K M and Lin J Q 2020 *Chin. Phys. B* **29** 124209
[27] Daniel D, Alejandro M and Hahn D W 2019 *Appl. Spectrosc.* **74** 42
[28] Sitangshu C, Manjeet S, Bishnu P B, Uday K S, Suraj P and Arnab S 2019 *Anal. Bioanal. Chem.* **411** 2855
[29] Li W T, Zhu Y N, Li X, Hao Z Q, Guo L B, Li X Y, Zeng X Y and Lu Y F 2018 *J. Anal. At. Spectrom.* **33** 461
[30] Pathak A K, Singh A, Kumar R and Rai A K 2018 *Natl. Acad. Sci. Lett.* **42** 87
[31] Liu X D, Feng X P and He Y 2019 *Renew. Energ.* **143** 176
[32] Lu Y Y, Boukharouba K, Boonaert J, Fleury A and Lecoeuche S 2014 *Neurocomputing* **126** 132
[33] Jia J W, Fu H B, Hou Z Y, Wang D H, Ni Z B and Dong F Z 2018 *Plasma Sci. Technol.* **21** 034003
[34] Wang Q Q, Teng G, Li C Y, Zhao Y and Peng Z 2019 *J. Hazard. Mater.* **369** 423
[35] Chen X, Li X H, Yu X, Chen D Y and Liu A C 2018 *Spectrochim. Acta, Part B* **139** 63
[36] Lu X N, Deng L, Du J C and Vienna J D 2020 *J Non Cryst Solids* **553** 120490
[37] Guo L B, Hao R F, Hao Z Q, Li K H and Zeng X Y 2013 *Acta Phys. Sin.* **62** 224211 (in Chinese)
[38] Jin J, Chen Y, Pei L S, Hu C J, Ma X X and Chen C X 2000 *Acta Phys. Sin.* **49** 1689 (in Chinese)
[39] Couairon A and Mysyrowicz A 2007 *Phys. Rep.* **441** 47
[40] Li S Y, Guo F M, Song Y, Chen A M and Jin M X 2014 *Phys. Rev. A* **89** 023809
[41] Anabitarte F, Mirapeix J, Portilla O, Lopez-Higuera J and Cobo A 2012 *IEEE. Sens. J.* **12** 64
[42] Campanella B, Grifoni E, Legnaioli S, Lorenzetti G, Pagnotta S, Sorrentino F and Palleschi V 2017 *Spectrochim. Acta, Part B* **134** 52
[43] Liu W, Théberge F, Arévalo E, Gravel J F and Chin S L 2005 *Opt. Lett.* **30** 2602
[44] Valenzuela A, Munson C, Porwitzky A, Weidman M and Richardson M 2014 *Appl. Phys. B* **116** 485
[45] Zhu X Q, Xu T, Lin Q Y, Liang L, Niu G H, Lai H J, Xu M J, Wang X, Li H and Duan Y X 2014 *Spectrochim. Acta, Part B* **93** 8